DOCUMENT RESUME

ED 106 362 TM 004 478

AUTHOR	Sharon, Amiel T.
TITLE	Technical and Practical Problems in Assessing Experiential Learning.
PUB DATE	[Apr 75]
NOTE	10p.; Paper presented at a joint session of the Annual Meeting of the American Educational Research Association and the National Council on Measurement in Education (Washington, D.C., April 2, 1975)
EDRS PRICE	MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS	College Credits; Educational Innovation; Educational Quality; *Learning Experience; Learning Processes; Measurement Techniques; Relevance (Education); *Student Evaluation; Teacher Evaluation; Test Bias; *Testing Problems; *Test Reliability; *Test Validity
IDENTIFIERS	*Experiential Learning; Off Campus Experiences

ABSTRACT

The measurement difficulties of experiential learning arise because assessment is often individualized, off-campus learning experiences are not well structured or defined and paper-and-pencil tests are often inappropriate measurement devices. The assessment process requires quality control to assure that it can accurately indicate whether learning has taken place and to verify that the learning meets an educational standard of quality. Special attention must be given to the reliability and validity of the assessment procedures. Of special significance is the notion of standards. Both the norm-referenced and criterion-referenced approaches can be employed to determine whether the student has achieved an acceptable level of learning. (Author)

ED106362

Technical and Practical Problems in Assessing Experiential Learning[1]

Amiel T. Sharon

Educational Testing Service

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Since experiential learning involves individuals who learn in a variety of ways in off-campus settings, there are a number of special technical and practical problems in its measurement and evaluation. Some of the difficulties in measuring nontraditional education are shared with the assessment of conventional classroom achievement while other psychometric problems are unique to experiential learning. Measurement difficulties arise because assessment is often individualized, off-campus learning experiences are not well structured or defined and traditional paper-and-pencil tests are often inappropriate measurement devices.

One of the functions of the assessment of experiential learning is to maintain quality control and support the credibility of nontraditional educational programs. The assessment process itself, however, requires quality control to assure that it can accurately indicate whether learning has taken place and to verify that the learning meets an educational standard of quality. The concepts of reliability and validity which are so important in the assessment of human abilities and educational achievement are equally important in the evaluation of off-campus experiences. Of special significance is the notion of standards. Since life experiences

---

[1]This Paper was presented at a joint session of the American Educational Research Association and the National Council for Measurement in Education, Washington, D.C., April 2, 1975.

TM 004 478

are unique to each individual, the establishment of standards by the traditional normative approach is often inappropriate. Finally, since paper-and-pencil tests are often irrelevant for the measurement of "learning by doing," the use of alternative assessment procedures raises questions of cost and practicality.

Reliability

Reliability is an indication of the consistency or stability of measurement results, whether they are in quantitative or qualitative form. Various errors cause even the best possible measurement device to be only an imperfect measure of the student's "true" knowledge or achievement. As Stanley (1971) points out, however, there is no single universal reliability for a test because different sources of variance of scores on a particular test can be allocated differently to true variance and to error variance. Stanley has defined some six different categories of score variance which may or may not allocate to error. These categories include combinations of lasting, temporary, general and specific characteristics of examinees as well as systematic or chance factors and variance not otherwise accounted for.

The factors falling primarily within the category "systematic or chance factors affecting the administration of the test or the appraisal of test performance" are almost always treated as error in the measurement of educational achievement. These factors include the conditions of testing, interaction of examiner-examinee traits and bias in grading or rating

performance. These sources of variation are of greater concern in the assessment of experiential learning than in the evaluation of the typical classroom learning. Since written objective examinations used in the classroom are administered and scored under highly uniform conditions, the error attributable to systematic or chance factors which affect the administration and scoring of such tests is negligible. When assessment is conducted on a one-to-one basis, however, the chance of random as well as systematic error is likely to increase and unreliability becomes a greater concern.

Most of the techniques used in assessing experiential learning require the judgement of a human observer or evaluator. These techniques include interviews, performance tests, simulations, ratings of performance and ratings of products. A major source of error, and therefore unreliability, in all of these assessment techniques is therefore the human observer who records, quantifies and evaluates the performance and the products of the student.

The interview technique which is probably the most widely used in assessing off-campus experiences is also the one most susceptible to measurement error and therefore to unreliability. Both the interviewer and the interviewee are a source of this error. The accuracy of the information given by a student during an interview will vary according both the physical and psychological environment in which the interview takes place (e.g. the "tone" of the interviewer), the recency of the student's learning experiences, and the clarity and relevancy of the questions asked. The reli-

ability will also be affected by the way the interviewer records, transcribes, codes, summarizes and evaluates the information provided by the student. An interview is likely to be reliable to the extent that the judgement of two or more interviewers are similar on the nature and level of a student's learning.

The reliability of assessment by the use of simulations, performance tests and direct observation and evaluation of performance or products depends largely on the accuracy of the judgement of the rater or evaluator, the kind of behavior being rated, and the rating method used. Typical biases of raters which tend to reduce accuracy are the tendencies to rate too leniently or severely, to allow an outstanding or inferior trait or aspect of performance to influence the rating of other traits, to judge students according to a personal stereotype or to prejudge on the basis of an initial impression.

How can the reliability of assessing experiential learning be improved? Since the interviewer, the rater, or the evaluator account for much of the error variance and unreliability, it is reasonable to assume that more accurate ratings will result if the person doing the assessment is trained in the interview and rating process and devotes a sufficient amount of time to assessment. Just as lengthening a typical written objective examination will increase reliability, increasing the number observations of the student in a real or simulated situation is likely to provide a more accurate indication of his competence. Finally, the more persons rate an individual, the more reliable are the pooled ratings likely to be if each rater is familiar with the individual and the performance being rated.

Validity

The validity of an assessment technique concerns what the technique measures and how well it does so. The question of what is being measured is especially important to experiential learning because it relates to the content validity of the assessment process. There is a need to specify intended outcomes of experiential learning programs in terms of measurable competencies and how these outcomes relate to formal classroom learning. Some institutions, for example, desire to measure off-campus learning which is equivalent to classroom learning while others have major interest in measuring those competencies which complement rather than duplicate the ones gained in formal courses. The specific learning outcomes for the individual student may be specified both by the institution, as represented by a faculty advisor or mentor, and by the student. Regardless of who defines the objectives, they must be spelled out in fairly specific terms. This specificity is necessary in order to be able to develop content-valid assessment techniques and to help the student understand what will be expected of him.

The validity of different procedures for assessing experiential learning will depend on the types of skills measured, the technical quality of the procedure used, the way by which the procedure is administered and the manner in which the measurement results are interpreted. No single type of assessment is likely to be superior to all others as each has certain disadvantages in specific situations. Furthermore, decisions concerning placement and credit should not ordinarily rest on a single technique or measure of competence. Multiple measures are likely to provide added

comprehensiveness and validity to the entire assessment process.

Many of the techniques that are considered for assessment of experiential learning can be categorized in terms of similarity between the learning and the assessment situation. These range from those that have a strong element of realism (for example, performance tests) to those that present a highly artificial and restricted assessment situation (such as written objective examinations). Performance tests or work samples usually attempt to reproduce all or an important part of the actual operations and tasks of a job. Their validity is likely to be in direct proportion to how well they actually reproduce the work involved in a job. They are most appropriate for measurement of work experience in vocational-technical fields as well as in foreign languages, art and sports. Simulations are somewhat more removed from reality than are performance tests. They often concentrate on the measurement of attitudes, motivation, social skills, leadership ability and decision making skills as opposed to more simple and directly observable skills. Their validity is dependent, to a large degree, on the extent to which the essential elements of a particular job are reproduced. Paper-and-pencil examinations often cannot present a situation that resembles the real one. Nevertheless, they have content validity for the measurement of competencies such as creative writing, proper use of grammar and knowledge of definitions. Written tests can usually sample a wide range of a subject-matter and can be scored with relatively high and even complete objectivity.

It is likely to be more difficult to demonstrate the empirical than the content validity of techniques used to assess experiential learning. Since the number of students within a given institution who are assessed for experiential learning is generally small and these students' experiences are highly varied, it would be difficult to identify a sufficiently large sample for an adequate validity study. In addition, relevant criteria against which to validate the measurement techniques are frequently not available.

Standards

It is not enough to measure experiential learning--there is a need to decide whether the quality and extent of the learning is at an acceptable level in order to aid in making decisions concerning the amount and kind of academic credit to be awarded to the student. An important concept in the assessment process, but one about which there is a lack of clarity, is that of standards. Often this concept is referred to incorrectly as consistency. (It was pointed out earlier that consistency refers to reliability.) A standard is anything taken by general consent as a basis for comparison. Thus, a student falling below a general standard in reading would generally be regarded as a poor reader.

There are two basic approaches to the development of standards: the norm-referenced and criterion-referenced approaches. The norm referenced approach involves a comparison of the learning of the student with that of a well-defined norms group. The performance of a group on a specific measure defines the standard. The performance of the student

on the same measure determines whether he is above or below the standard. This approach can be used with students who have experience which can be directly related to a specific course. If valid measures of course outcomes are available, such as an end-of-course achievement examination, then such examinations could be administered to students completing the course in order to define the standard or norm. Nontraditional students who can score as well as the average student with a passing grade in the course can be exempted from the course and awarded credit for it. Since off-campus learning experiences are generally highly unique, the norm-referenced approach is likely to have only limited applicability.

The criterion-referenced approach to setting standards involves the use of expert judgement as to what constitutes satisfactory performance. It involves specifying what a student must know or what he must be able to do in order to demonstrate his competence in a particular subject. The responsibility for specifying the standards often rests with both the institution and student as is the judgement whether the standards have been met. Many institutions make use of faculty and external experts to review student portfolios and to interview students regarding their experiences. These subject-matter experts bring their collective judgements to bear on the quality and level of learning acquired in off-campus settings and its relevance to the student's and to institution's goals.

Conclusion

The foregoing technical problems must be considered in the larger context of the measurement and evaluation of experiential learning. The

practical aspects of assessment such as cost, faculty involvement, the amount and nature of credit to be granted, and the transferability of the credit cannot be divorced from the technical aspects of assessment. High quality assessment from the technical standpoint is feasible only if it is economically feasible. The success of process will also depend on the capability and willingness of the college faculty to invest the time and effort necessary to conduct individualized evaluation. Nevertheless, the psychometric aspects of assessment are likely to be as important, if not more so, than the administrative aspects. They are at the heart of the larger issues in assessment--whether it is fair and equitable to students, whether it is credible to third parties and whether it can be successfully used to integrate students' life experiences with formal higher education.

REFERENCE

Stanley, Julian C. Reliability. In R. L. Thorndike (Ed.) Educational Measurement. Washington: American Council on Education, 1971.